DOCUMENT RESUME

ED 160 649 TM 007 881

AUTHOR Bank, Adrianne; Burry, James

TITLE Directions in Testing and Evaluation: CSE, 1978.

INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.

PUB DATE Jul 78
NOTE 13p.
AVAILABLE FROM James Burry, Managing Editor, Evaluation Comment, 145 Moore Hall, University of California, Los Angeles, California 90024 (Free)
JOURNAL CIT Evaluation Comment; v5 n3 Jul 1978

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Educational Practice; *Educational Research; Educational Researchers; *Educational Testing; *Evaluation; Evaluation Methods; *Program Descriptions; Publications; Staff Utilization
IDENTIFIERS *Center for the Study of Evaluation CA

ABSTRACT

A number of current activities at the Center for the Study of Evaluation (CSE) are described, including research programs, evaluation services, workshops and field training programs, publications, and information dissemination. The Evaluation and Decision Making Program studies evaluation from three different perspectives: (1) the program level, which examines the activities of program directors and evaluators; (2) the school district level, which integrates field work and surveys; and (3) the philosophical level, which examines the basic issues and theories of evaluation. CSE's Measurement and Methodology Program currently includes four projects: the design and use of tests; multilevel analysis; psychometric problems in achievement tests; and reanalysis of large-scale data sets. The Bilingual Evaluation Program emphasizes research on the characteristics of an appropriate transition to English-language instruction; technical assistance to evaluators; and training of female and minority group evaluators. The list of new or recent CSE publications includes monographs, reference materials about tests, and kits for assessing reading instruction. A list of CSE personnel is also included. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Center for the Study of Evaluation

CSE

EVALUATION COMMENT

UCLA

July 1978 Vol.5, No.3

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED160649

TM007881

In this issue:

Evaluation Comment resumes publication after a long quiet period. In the interim our many friends have expressed concern over the future of *Evaluation Comment*. We want to reassure you that you remain on our mailing list, and that we plan to maintain regular publication. We will have three or four issues each year. Each will have a CSE senior staff member as guest editor. Future topics include: proceedings of CSE's 1978 Measurement & Methodology Conference; evaluation of bilingual programs; issues in the training of evaluators. We would welcome your suggestions for other topics.

The purpose of this issue is to inform you of the full range of current activities at CSE. Our intensified research efforts emphasize two areas we believe important to education today: research into testing practices and their interpretation; and research into evaluation practices and their utilization. We continue to explore these subjects with specific relation to bilingual education.

CSE's Visiting Scholars Program brings to the Center distinguished experts from universities, school districts, and state education agencies who not only deepen and broaden our current work but also point to directions for the future.

Likewise, our annual and special conferences on measurement and on evaluation assist us in our agenda-setting endeavors. We have several books in press, three additions to our monograph series, and many new entries into CSE Reports.

We describe for you here our three major research programs along with our field service and dissemination activities (for the convenience of those who skim, the general descriptions are set in bold-faced type).

Please let us hear from you. Comments on our *Comment* will be appreciated.

Directions in Testing and Evaluation: CSE, 1978

Adrianne Bank
James Burry

The Center

The work of the Center for the Study of Evaluation is devoted to the theory and practice of evaluation and to the supporting areas of measurement and methodology. Our three major programs are concerned with *Evaluation and Decision Making*, *Measurement and Methodology*, and *Bilingual Evaluation*. In addition, we maintain two units—*Evaluation Field Services* and *Dissemination*—which assist in the utilization of research from the three programs.

Since our inception in 1966, we have studied evaluation and measurement across the entire spectrum of conceptualization, research, development, training, and dissemination. We have also conducted evaluations which have methodological challenge or implications for policy.

CSE's prime mission is to study evaluation as it affects the formulation, development, and delivery of educational programs and services. Our commitment to evaluation grows from the belief that school practices and the competencies and satisfactions of those who participate in the educational enterprise can benefit from information collected in accordance with social science methodologies. But we do not wish to perpetuate the simplistic notion that more educational evaluation is what is needed; possibly education needs fewer but more carefully delineated evaluations that are designed and reported in ways that take into consideration the nature of the audience and their power to effect alternative actions. Our work is intended to identify the critical attributes of such evaluations and the milieu in which they are conducted.

Public recognition of the field of evaluation is little more than ten years old. These ten years saw exploratory model building and the development of competing approaches to evaluation; early evaluation reflected the rapture with linear research and development models that characterized the larger educational enterprise. For example, dominant methodology in early evaluation was extrapolated from experimental design; evaluation imitated research. But evaluation has been evolving. Recognition of systems analysis procedures influenced much of later evaluation precept and practice. In turn, these procedures were superseded, in part, by others that emphasized more qualitative ways of understanding the effects of educational and social action programs. But the particular benefits of each of these orientations are not now in question; rather, the issue is one of choosing among desires for richness of data, providing texture for interpretation, and the need for inference from more precise data to a narrower band of phenomena. Instead of fitting data to a model, we prefer to fit an evaluation model to the circumstances.

CSE's location on the UCLA campus facilitates our interest in the particular needs of a metropolitan, multilingual community. Our position within the University of California system and the Graduate School of Education allows us to draw upon expertise from the University at large, as well as upon the School's well-recognized faculty, for support and participation.

Each project within a CSE program is headed by a senior professional Project Director. Project Directors are a highly trained group reflecting a variety of experiences and multidisciplinary backgrounds. They are almost equally divided between individuals who are concerned with technical issues associated with evaluation, measurement, design, and analysis, and those who are particularly concerned with the improvement of educational services and the planning of instruction. Virtually all have had teaching experience. Each project may have one or more research associates, usually advanced graduate students.

Faculty members from the School of Education provide expertise in the planning and conduct of project work. Faculty participation falls within one of two categories. The closest faculty relationship is that of Faculty Associates, who perform specific tasks in relation to CSE work and provide general advice, assistance, and review for a number of projects. In addition, faculty members may serve as Faculty Affiliates who either contribute regular proportions of their time as service to the University, or participate with CSE as needs and interests warrant. These faculty members represent such disciplines as Law, Medicine, Management, Architecture and Urban Planning, Political Science, Psychology, and Philosophy, in addition to Education.

The senior professionals, associates, faculty, and support staff provide experience and continuity for our efforts; their talents are supplemented by Visiting Scholars—from universities and school districts within the United States as well as from abroad—who study and work at CSE for periods of time ranging from several weeks to a year.

At CSE, we are continuing to study and improve the quality and utility of evaluation. In doing so, we take the role of national consumer advocate in the area of evaluation. Our advocacy is in the interest of the consumers of educational services—the students, the supporters of the educational enterprise, the professional and lay communities. Our goal is to raise the standards for producing evaluations so that they are accurate, credible, and useful. This goal is reflected in our three current research programs—the Evaluation and Decision Making Program, the Measurement and Methodology Program, the Bilingual Evaluation Program—and in our Field Services and Dissemination units.

Center for the Study of Evaluation
UCLA

Eva L. Baker
Director

EVALUATION COMMENT is published by the Center for the Study of Evaluation, University of California, Los Angeles. CSE is the only organized federally-supported agency devoting its total scope of work to the exploration and refinement of strategies in evaluation. CSE was established by the U.S. Office of Education in 1966 and is currently supported by the National Institute of Education. For 12 years, CSE has furnished the educational community with products and research which meet the increasing requirement for the evaluation of educational and social action programs.

Each issue of **Evaluation Comment** will discuss topics in educational evaluation by presenting articles on evaluation theory, procedures, methodologies, or practices. For each issue of **Comment**, CSE will request contributions on a specific topic in evaluation from recognized experts in the field; unsolicited manuscripts will not be accepted.

Evaluation Comment has a current readership of over 10,000 readers. Each person on our mailing list receives a free copy of **Comment**. Where additional copies are needed readers are encouraged to reproduce **Comment** themselves. To be placed on our mailing list, please write to:

James Burry, Managing Editor
Evaluation Comment
145 Moore Hall
University of California, Los Angeles
Los Angeles, California 90024

Editorial Board

Eva L. Baker	Chairperson
Tomas Arciniega	Dean, School of Education San Diego State University
William Johnston	Superintendent, Los Angeles Unified School District
John B. Peper	Assistant Superintendent for Planning, Milwaukee Public Schools
Michael Scriven	Director, Evaluation Institute University of California, San Francisco
Robert E. Stake	Director, CIRCE, and Professor of Educational Psychology at the University of Illinois
Daniel L. Stufflebeam	Professor of Educational Leadership and Director of the Evaluation Center, Western Michigan University

Administrative Staff

Eva L. Baker, Director
Adrianne Bank, Associate Director
Richard Littlestone, Fiscal Officer
David Henderson, Business Manager
Marlene Henerson, Coordinator of Dissemination
Jane Coogan Beer, Director, Public Information
James Burry, Director, Field Services

Senior Research Staff

Frank Capell
Kathleen B. Fischer
Joan Herman
Ruddie Irizarry
Betty Johnson
Catherine Dillon Lyon
Lynn Lyons Morris
Edys Quellmalz
Clinton B. Walker
Rand R. Wilcox
Jennie Yeh

Faculty Associates

Marvin C. Alkin	Charles O'Reilly
Beatriz M. Arias	David O'Shea
Leigh Burstein	Harry F. Silberman
Frederick Ellett, Jr.	Concepcion Valadez
Oscar Grusky	Richard C. Williams
	Lynne G. Zucker

Recent Visiting Scholars

Larry Barber	Assistant Superintendent for Research, Development, and Evaluation, Eugene Public Schools; Eugene, Oregon
Gene V Glass	Professor, University of Colorado; Co-Director of Laboratory of Educational Research
Egon G. Guba	Professor, University of Indiana; Associate Dean for Academic Affairs
James Jacobs	Superintendent, Cincinnati Public Schools
Michael H. Kean	Executive Director for Research and Evaluation, Philadelphia School District
Alex Law	Director, Office of Program Evaluation and Research, California State Department of Education
Arthur Lumsdaine	Professor, University of Washington
Bernadine Stake	Research Assistant, University of Illinois; Alderwoman, Urbana City Council
Robert Stake	Professor, University of Illinois; Director, Center for Instructional Research and Curriculum Evaluation
Howard Sullivan	Professor, University of Arizona

The Evaluation and Decision Making Program

Evaluation has grown faster than our abilities to document or direct it. As part of our continuing efforts to make sense of the various ideas and activities calling themselves evaluation, we are working in three selected areas. These areas do not deal with the entire terrain of evaluation nor do they fit neatly into an overall map of the territory. Rather, they have grown out of our past experiences and appear to have important implications for the field.

The three projects in this program examine the current condition of evaluation from three different points of view: organizational, interpersonal, and analytic.

The project on Evaluation and Decision Making in School Districts looks at evaluation from the organizational perspective. Evaluation activities are carried out in school districts in a variety of ways. One of the most common organizational arrangements, especially in large districts, is the research, development, and evaluation unit located within the administrative structure of the district. These units have evolved independently of one another, each responsive to the particular historical and structural context of its district. For the past year, we have been studying these units in an attempt to understand their development and functioning, as well as to note their regularities and peculiarities.

The project on Evaluation and Decision Making at the Program Level looks at evaluation from the interpersonal perspective; it examines the ways in which the evaluator and the program director deal with one another. These roles, evaluator and program director, cannot be divorced from their institutional context, but they can be examined for the variables which seem to explain or predict the effectiveness of the evaluation.

The third project, Philosophic Inquiry into Educational Evaluation, looks at evaluation from the analytic perspective. It applies ways of thinking used by philosophers, such as paradigm cases, metaphor, and application of logical and ethical criteria, to conceptualizations of evaluation.

Evaluation and Decision Making in School Districts

This project analyzes educational evaluation at the local educational agency (LEA) level. Our research methodology is grounded in integrated field work and survey activities.

One major question drives the project: To what extent are school district evaluation offices contributing to the renewal of public education? Previous CSE work reveals a major information gap in educational evaluation: we simply do not know which school districts have centralized program evaluation units, what they are doing, or how evaluation processes and products are used by the "stakeholders" in public education.

Through a recent survey, we identified those school districts with centralized organizations responsible for program evaluation. Every school district in the country with enrollment of 10,000 plus (n = 750) was surveyed; 335 centralized program evaluation units were identified. A fifty-percent sample of districts enrolling 5,000 to 9,999 students (n = 573) was also surveyed; 74 centralized program evaluation units were identified. A total of 409 school districts with centralized program evaluation units were thus identified throughout the United States. The survey of these units follows several areas of inquiry:

- What are the functions of offices of evaluation?
- How are evaluation unit processes, personnel, and products utilized?
- Who uses evaluation information?
- To what extent do these participants in public education use the information?
- How are evaluation offices organized?
- What are the characteristics of evaluation personnel?
- What resource constraints and requests are reported by heads of evaluation organizations?
- What characteristics of school districts affect evaluation offices?

While the surveys of school districts and evaluation unit heads were being designed, three case studies of school district RD&E operations were completed. The same CSE team visited three school districts in different parts of the country (Northwest, East, and Midwest) with enrollments of 20,000, 260,000, and 65,000. Because of the planned informational overlap between the field work activities and the survey, both are guided by the same areas of inquiry. Survey results will enjoy some interpretive advantages from the field work; in turn, the survey results will lead to improved post-survey field work.

We are also examining a sample of recent evaluation reports requested from each survey respondent. The purpose of this analysis is to describe actual evaluation unit activities; classification of the documents will be based on the same unit activities that are central to the survey. For example, do reports emphasize data collection, testing, development, program evaluation, or research?

Evaluation and Decision Making at the Program Level

Evaluation requirements for educational programs have reflected a hope that evaluation information would improve the

quality of decision making, that it would contribute to the rationality, objectivity, and defensibility of decisions.

These hopes as yet seem unfulfilled; evaluation seems to have suffered from under-utilization. But several empirical studies of this problem suggest that evaluation may influence decision making. These studies suggest two important points. First, our definition of evaluation utilization may need to be broadened beyond the restrictive notion that evaluation is utilized only if it decisively determines some milestone decision; our conceptualization of evaluation should encompass the more subtle and incremental kinds of influences that evaluation can have on programs. Second, the key to increasing evaluation utilization may lie in the careful study of evaluation practice. All over the country school people are trying to conduct good evaluations and make intelligent use of evaluation information. If we can identify their strategies, successful and unsuccessful, we may learn a good deal about making evaluation more useful.

This project is studying evaluation's impact at the program level to determine the extent to which evaluation *does* influence program-level decision making, and to identify the constraints on evaluation impact and the conditions that might encourage higher levels of evaluation utilization. We have conducted five exploratory case studies of program evaluation in schools. Each case study focused on a single instance of program evaluation and attempted to identify the influence that evaluation information had on program decision making. The end product of each was a detailed case study report describing the program, its evaluation, and our findings on utilization.

We use a retrospective approach in the case studies. This involves in-depth interviewing of the operational staff and the program evaluator and, as supplement to the interviews, reviewing documents such as program proposals and evaluation reports. In the interviews we avoid being directive; we do not want to impose biases that could lead the informants to tell us the story as we might want to hear it rather than as they would naturally present it.

We also look for the operation of such factors as the social and organizational context of the program; the characteristics of the evaluator, program staff, and program administrators; the purposes and procedures of the evaluation; the interactions among the major actors; the manner in which evaluation findings are communicated.

Interviews with core informants frequently lead us to other interviewees—an influential teacher, principal, or counselor. These newly identified informants may in turn lead us to others; the interviews snowball until all the major actors in the evaluation situation appear to be identified and interviewed.

As a screen against error, the first draft of the study is reviewed by the study *team* and revised to answer questions raised by the team. The draft is next circulated to the interviewees for comment and criticism. After each key informant has examined the draft, we schedule a personal interview to discuss the report. Using the information gained in these sessions, the final report is prepared by the team. Informants are then invited to submit remarks on the case study findings, the case study process, and the study of evaluation and evaluation utilization. Their remarks are included in the final report.

Our findings so far are that none of the evaluations we have studied have had a "make or break" impact on the programs. However, the relationship between the evaluator and the program director seems to be of considerable importance. The evaluator who is able to focus on the decision needs of the program director seems more likely to create an influential evaluation. But there is often a tension between trying to serve the program director and fulfilling the mandated evaluation role. The evaluator who concentrates exclusively on the mandated tasks may lose the local audience; the most successful evaluators seem to be those who carry out the required tasks and at the same time invest great effort to help local program people with the day-to-day decisions that cumulatively make a difference.

Later in the project, we will conduct a series of confirmation studies to verify our findings and to provide further insights into the key characteristics and their interaction and impact on utilization.

Philosophic Inquiry into Educational Evaluation

Evaluation does not have a generally agreed-upon philosophic basis; the many practices called evaluation rest on assumptions, often implicit and sometimes at odds with one another. We are trying to bring clarity to this area by examining some basic issues in evaluation and developing a series of essays on these issues. Practice without good theory is blind; theory without practical import is useless.

For example, one of our early papers examines the foundation of educational evaluation, and draws key distinctions between empirical and normative theorizing and between the various forms of educational evaluation. An empirical theory describes what is the case; a normative theory prescribes what the evaluator should do. We argue that theories of evaluation are founded on both empirical and rational-ethical grounds.

A second paper presents a philosophical investigation of the theories of needs assessment. Almost all theories of educational evaluation recognize that the determination of the proper standards, criteria, or values by which the program can be judged is centrally important to performing evaluations. This paper assesses these theories of needs assessment in terms of their ethical and practical implications and commitments. We contend that the theories of needs assessment have not been justified on rational or ethical grounds.

Among the issues that we are presently examining are:

Evaluating Teacher Effectiveness. Educational evaluation is centrally interested in evaluating teaching, learning, and curricular programs. This essay will investigate, from a philosophical perspective, the theory and practice of evaluations of teacher effectiveness. Our primary concern is to explicate the conceptual and methodological presuppositions of current strategies used to assess teacher effectiveness and to present arguments for and against the plausibility of the various presuppositions. Such work has immediate implications for teacher effectiveness research.

Educational Evaluation and Ethical Considerations. This essay will look at the presuppositions that concern the philosophy of social science and ethics. Ethics is, in part, interested in the systematization and clarification of views about what is right or wrong, good or bad, just and unjust, obligatory or permissible. The essay will consider (1) whether moral (or ethical) judgments should be used in evaluation, or whether evaluations should employ only descriptive items; and (2) the ethical presuppositions of the dominant views of educational evaluation.

The Point and Performance of Evaluation. This paper will criticize the alternative practices underlying educational evaluation. In addition, we will put forward a conception of evaluation that is justified on rational and ethical grounds. This final essay, drawing on the recent work of philosophers in both the Anglo-American and continental traditions, will argue that evaluation should be thought of as practical reasoning, and will show the advantages and limitations of such a view and defend it against the current, dominant views.

In Preparation . . . a book on Evaluation Utilization in Local Schools: A Case Study Approach.

The Measurement and Methodology Program

The measurement of academic achievement rests at a strange point in the current national discussion of educational issues. On the one hand, scholars, often supported by teachers, have expressed concerns about the quality and utility of achievement tests. Yet the community, represented by local school boards, state legislatures, and other decision-making groups, is placing increasing reliance on testing to assess the effects of educational efforts. Public interest in testing is seen in recent calls for the measurement of learner proficiency with attendant suggestions that "no-performance, no-promotion" strategies be adopted and that high school minimal standards be set. There is also contention about which forms of tests are most appropriate for various purposes. Standardized achievement tests, while permitting comparisons among groups and individuals, are criticized as being unresponsive to particular program efforts or unfair to certain sets of students. But in turn, the newer forms of achievement tests, so-called competency-based or criterion-referenced tests, are faulted for their frequently narrow focus and lack of comparative norms. The necessity for developing measures and their supporting statistical procedures in which we can place our confidence is obvious.

Our past work has a bearing on this program. The development of the Instructional Objectives Exchange, the System for Objectives-Based Assessment in Reading, and the continued research on bilingual assessment models provide instances where recognition of the needs of the field for improved measurement procedures generally outstripped the methodological base for the developments. We have devoted effort to the preparation and application of criteria for the evaluation of commercially available tests because of our recognition of the high level of public expenditure in the testing area. We have evaluated norm-referenced tests for pre-school, elementary, and secondary students, covering basic and higher order intellectual skills as well as the affective domain. We have continued this work into the evaluation of commercially available criterion-referenced tests. The findings from these efforts document the exceedingly uneven level of quality in available test materials; the overall level of understanding of tests, their design, their interpretation, and their usefulness for classroom and policy decision making is at best rudimentary.

Measurement theory has relatively strong psychometric bases but lacks subject matter or content understanding. Furthermore, the substantive theory linked to quantitative concerns draws on psychological theory, primarily in the areas of intelligence and personality. We do not think this traditional psychometric approach is necessarily appropriate for achievement tests. A theory of test *content* with corresponding *qualitative* and *quantitative* modes needs to be explored. Without such a theory, systematic improvement of the quality of achievement tests is unlikely. Researchers in measurement theory seem to limit their study to decision-theoretic approaches and questions of test reliability using alternative statistical models. This kind of inquiry is extremely important and will be continued in the program. However, relatively little work has been conducted in other aspects of theory development, such as the design characteristics of tests.

To remedy this need, our program in measurement and methodology intends to develop an integrated theory of achievement testing. This theory, minimally, should address such concerns as: (1) conceptual and empirical analyses of alternative measurement approaches; (2) development of models and procedures for the analysis of test quality such as reliability and sensitivity through the use of statistical and other indicators; and (3) modeling and experimentation in analysis procedures appropriate for the interpretation of test results for various student populations in various environments.

Currently there are four projects working in this program:

The project on Design and Use of Tests is working on test design procedures primarily in the area of domain-referenced testing.

The Multilevel Analysis project is looking at: (a) alternative quantitative approaches for analyzing multilevel data on educational effects; (b) the properties of alternative measures, e.g., means, standard deviations, deciles, slopes of group-level (school, class) educational outcomes; and (c) the implications of other multilevel analysis work for test development and interpretation.

The project on Psychometric Problems in the interpretation and application of criterion-referenced and mastery achievement tests focuses on theoretical and statistical methodology for interpreting and applying criterion-referenced measurements.

The Reanalysis project is examining the sensitivity of a variety of non-standard quantitative methodologies for the detection of the effects of large-scale educational interventions.

The Design and Use of Tests

This project is conducting research on alternative test design procedures as they relate to different conceptions of the testing process. We are concerned with the general area of competency-based testing, specifically focusing on domain-referenced achievement tests. Domain-referenced test design techniques are based on the assumption that by clarifying content and criteria we may develop a set of test items that representatively sample an individual's or group's proficiency with regard to a given domain. Ensuring that items which test an instructional outcome properly sample the domain of that outcome has been the topic of much recent work in testing.

We wish to explore the benefits of clearly understandable domain-referenced test specifications within the reality of schooling. To that end, we have established criteria for good tests that will guide us in the selection of specific research activities. These criteria consist of (1) *publicness*: Does the test include a clear description of its design and rules for item generation? (2) *economy*: Does the test require the minimal student time consistent with the need to make a fair judgment of progress? (3) *instructional relationships*: Does the test measure dimensions of student competence that are amenable to instructional manipulation? (4) *meaningfulness*: Does the test assess non-trivial accomplishments that relate to lifelong competencies?

We are also generating information about interventions that will help teachers make the most effective use of test information. For example, we are working with teachers to find out what kinds of tests they use, how they select them, how often they administer them, and the extent of teacher knowledge of the technical aspects of testing. As a corollary to these activities, we intend to look at the varying kinds of support materials provided to teachers in our effort to learn under what conditions the test, as an evaluation resource, can best assist teachers in instructional decision making.

A further concern in this project is content validity. Here we are using procedures developed in mathematics, particularly in the study of sets and groups. We are attempting to estimate the "belongingness" of an item to its domain or set. We are trying out an Item Rating Scale for achievement tests that we hope will help people to rate specific aspects of the content validity of particular test items.

In addition to our general work in test design and use, we are also looking at the specific skill area of writing. Our interest in

developing valid measures of writing competencies derives from inadequacies in existing approaches to composition assessment, and from the expectation that advances in the evaluation of written production will also contribute to methods for assessing student generated responses in other subject matter areas. Teachers and students both express interest in tests that directly assess skills likely to be useful (transfer) in future life activities. Thus, we are interested in writing since it is a skill necessary to "get out and get along" (high school exit, vocationally-related competency) and to "get in and go on" (college admission, college survival).

Among our major concerns in this part of the project are: comparing holistic, analytic schemes and multiple choice measurement in expository writing; validating scoring criteria across criterion groups in expository writing; examining the relationship of classroom instructional characteristics and writing in the descriptive/narrative mode; providing information about generalizability for writing skills across modes of discourse; and comparing criterion-referenced and norm-referenced tests of writing for various post-secondary populations.

In Preparation . . . a special report on teacher test use in schools.

Multilevel Analysis

The main premise guiding the Multilevel Analysis project is that educational researchers, evaluators, and decision makers need to understand that the analytical models they use to identify the effects of schooling on pupil performance influence their findings. Investigators who fail to consider the multilevel character (individual students, classes, schools, districts, states, etc.) of the educational system are likely to arrive at erroneous inferences about the impact of educational programs.

Many studies of educational effects have been analyzed at a single level (e.g., student, class, or school) and with data from more than one level (pupil-level outcome and background measures, school-level measures of educational characteristics). The Coleman Report, the IEA Six-Subject Surveys, and the evaluation of Project Follow-Through are just a few examples of large-scale studies in which questions concerning the proper levels of analysis arise. The problem is that in most cases the use of one level of analysis rather than another can lead to (a) different variables entering the analysis (b) spuriously large or small estimates of educational effects, and (c) inflated estimates of the effects of pupil background with corresponding decrements in the effects attributed to educational factors.

Correctly choosing among strategies for analyzing multilevel data is not a methodological decision independent of the substantive questions being addressed. Choices should be determined by the type of study one is conducting and the types of outcomes and processes under investigation. Even the casual observer of educational institutions recognizes the varied character and effects of educational processes. Different school districts have different resources to allocate among their schools and different guidelines governing allocations. Schools differ in their emphases, styles and quality of leadership, mixes of student backgrounds, and physical facilities. Teachers vary in personality, training, experience, attitudes toward teaching, preferences for kinds of children, curricular emphasis, style of teaching, and goals.

Given the many factors that enter into educational effects and given that these factors occur at multiple levels of the educational system, it is not hard to understand why the results of studies conducted exclusively at a single level of analysis have generally been discouraging. Different types of analyses of multilevel educational data should be used to address different questions. Typically research on schooling should ask questions answerable at multiple levels.

The multilevel character of educational data requires analytical models suited to the identification of educational effects at and within each level of the educational system. Our work in the Multilevel Analysis project examines the implications of insensitivity to the consequences of multilevel data; we are focusing on three methodological topics. First, several analytical models recently proposed for multilevel data are being investigated. Second, the properties of alternative measures (e.g., means, standard deviations, slopes) as indices of educational effects at the group (classroom, school) level are under study. Finally, the project will explore the implications from other aspects of its multilevel analysis work for test development and interpretation.

Psychometric Problems in Achievement Tests

This project has two general thrusts. The first is the further development and exploration of latent structure models designed specifically for achievement tests. The other is an investigation of the problems that occur when using item-sampling models.

We are extending latent structure models in two directions. First, we are looking toward a possible model for characterizing a population of examinees in terms of an item domain. Second, we intend to develop explicit estimates of a model's parameters for certain special cases. These solutions are important since there are technical difficulties with using iterative estimates of the parameters of these models; an iterative solution might not converge to the maximum likelihood estimates.

In a mastery test an examinee responds to a certain number of dichotomously scored test items. If the examinee's observed (number correct) score is greater than or equal to the passing score, a mastery decision is made and the examinee is advanced to the next level of instruction; otherwise a nonmastery decision is made and the examinee is given remedial work. One of the fundamental problems associated with mastery tests is determining an optimal passing score. To determine an optimal passing score, Huynh assumes that the true scores of the examinees follow a beta distribution and that there exists a function of the true scores which determines the probability of success on the external criterion. He also assumes that the conditional distribution of observed scores is given by the binomial probability function. Based on these assumptions, he first determines an optimal criterion score. If the true score is greater than or equal to the optimal criterion score, a mastery decision is called for. If the true score is less than the optimal criterion score, a nonmastery decision is desired. Once the optimal criterion score has been determined, it is possible, at least for some situations, to determine a passing score so as to minimize the risk or expected loss of the decision rule described above.

There is evidence that the beta-binomial model might give a reasonably good fit to mental test data. However, there are several reasons why alternative solutions need to be considered. First, the estimates of the parameters of the assumed beta prior used by Huynh might yield inadmissible (negative) values, in which case the procedure breaks down. Second, even for a moderately large number of examinees, the estimates of these parameters might be wildly inaccurate. Third, the beta-binomial model has the theoretical difficulty of allowing at most one mode for the true score distribution. Fourth, the assumption that the conditional distribution of observed scores is given by the binomial probability must be viewed as an oversimplification of the "true" situation when an item-sampling model applies.

In the project we have developed a procedure for determining the optimal passing score which makes no assumptions about

the conditional distribution of observed scores or the distribution of true scores. With this procedure it is possible to answer one of the two following questions: How many examinees are needed in order to determine an optimal passing score with probability at least P*? Given the test score results of N examinees, what is a lower bound on the probability of identifying an optimal passing score? When using Huynh's solution, it is not at all clear how certain we can be that an optimal passing score has been determined.

Reanalysis of Large-Scale Data Sets

This project is examining the extent to which newly proposed methodologies are sensitive to program effects, and the extent to which alternative strategies can elucidate the relationship between process and outcome. Among the questions we are considering are:

- Does consideration of alternative units of analysis and indicators of group-level outcomes lead to different conclusions about program effects and the relationships between school processes and educational outcomes?
- Does application of alternative techniques for setting proficiency standards alter the detection of program effects and/or the relationship between school processes and educational outcomes?
- Does application of alternative statistical methods result in different conclusions than those derived from methods such as analysis of variance and analysis of covariance?

In dealing with the questions of alternative units of analyses, we will reanalyze the relationships between process variables and outcome measures using pupils-within-classes and classes-within schools, as well as classes and schools as units of analysis. We believe that other characteristics of the distribution besides group means (e.g., the spread and shape of the distribution or the proportion of students above or below some standard) should be considered in data reanalysis. We plan to examine whether these alternative measures of outcomes are more sensitive to the purported effects of various educational programs.

Given the increasing interest in criterion-referenced measurement, there has been a corresponding increase in the development of psychometric technology for dealing with criterion-referenced and mastery tests that assess proficiency. One issue of particular concern is that of setting proficiency standards. In the project we will apply various standard setting techniques. The kinds of questions to be addressed include the degree to which assessments of program impact are consistent across the different techniques for setting proficiency standards, such as expert judgment, reference groups, and statistical decision theory.

Along with the increasing use of competency testing for partitioning performance of individuals into discrete categories, e.g., masters, nonmasters, and/or marginals, there is a growing need to identify methodologies appropriate to the type of binomial and multinomial data obtained from such testing. In evaluating program effects using mastery measures, there are several unique considerations (such as the shape of the distribution or the kind of outcome used) that should cause a reevaluation of the consequences of violating the assumptions required for performing the typically used inferential statistical techniques such as the analysis of variance or covariance.

One of the methods of analysis we plan to apply to the reanalysis of student outcome measures in large-scale data sets is the log-linear model. This model applies generally in the analysis of multi-factor, multi-response data when the response is qualitative but not ranked. The procedure for establishing the model matrix is straightforward and can easily be generalized to various experimental designs. The solution obtained through this method of estimation in many respects parallels the general linear model approach to the univariate and multivariate analysis of variance applied to continuous data.

In Preparation . . . the next in the CSE Test Evaluation Series: *Criterion-Referenced Test Evaluations*, providing descriptions and reviews of 63 criterion-referenced math and reading tests and test packages.

The Bilingual Evaluation Program

Since the late 1960s federal and state directives for bilingual education have created a variety of programs intended to meet the instructional needs of students for whom English is a second language. The ability of these programs to demonstrate effectiveness or to generate an evolving technology in bilingual instruction has been handicapped by, among other things, the lack of appropriate evaluation techniques and materials. The Bilingual Evaluation Program has addressed itself to these evaluation needs over the past years, and is continuing this work in three project areas.

The first of these areas is concerned with the problem of transition from bilingual programs to monolingual English instructional settings. Before students can be transferred out of a bilingual program, schools must document with reasonable certainty that such students are ready to comprehend fully and take advantage of instruction provided in English. The Bilingual Prediction Study is providing research-based information regarding appropriate configurations of language skills and other factors that will most likely ensure a child's success following transition to an English instructional setting.

The variety and discontinuity of bilingual programs hamper the interpretation of information regarding effectiveness; in the absence of such interpretation, it is difficult to recommend one instructional approach over another, to justify more expedient allocations of resources, or to recommend improvement in services. The Bilingual Evaluation Technical Assistance project (BETA) is developing workshop materials for practitioners and community members involved in bilingual education. These workshops present information in such areas as evaluation design and planning, measurement methodology, and analysis and interpretation of evaluation results. Through these workshops, we hope to meet the need for systematic methods and materials to aid bilingual educators in the difficult tasks of assessment and decision making.

The inclusion of a Women and Minorities Evaluation Training project as the third feature of this program seems particularly appropriate both in terms of the background of our personnel and the orientations of their projects. The Women and Minorities Evaluation Training project will try to define the needs—both technical and professional—existing among individuals who have evaluation responsibilities in schools or in the community. The research activities and the training materials to be developed within this program are designed to meet those needs.

The Bilingual Prediction Study

In the summer of 1975, the Department of Health, Education, and Welfare issued recommendations for meeting the requirements of the *Lau vs. Nichols* decision. Among the educational approaches that constitute appropriate "affirmative steps" at the

early elementary grades are bilingual/bicultural education programs and transitional bilingual education (TBE) programs. Both approaches incorporate instruction in the native language of the non-English-dominant student, while introducing English as a second language. Unlike bilingual/bicultural education programs, however, TBE's terminate language instruction "once the student is fully functional in the second language." Districts adopting the TBE approach, therefore, must provide predictive data which show that students are ready to make the transition from Spanish reading to English reading and will succeed in the content areas of the educational programs in which they are to be placed. Thus, if a district elects to accommodate the needs of non-English-dominant students through TBE's, it must furnish some assurance, based on valid and reliable evidence, that those students will profit meaningfully from their educational program once native language instruction is terminated.

The importance of making reasonably accurate estimates of potential performance cannot be exaggerated. For students for whom English is a second language, the consequences of premature or delayed transition can be harmful. Students who are transferred too early not only confront inordinate difficulties in trying to keep pace with instruction, but may experience damaging psychological effects as well. The situation produced by premature transition also fails to satisfy the fundamental legal mandate of "meaningful instruction." Similar harmful effects may be produced by unwarranted delay in transition.

The first phase of our research was devoted to a pilot study of Spanish-dominant third-graders in transition to fourth-grade English instructional settings. The sample consisted of a group of Mexican-American and native Mexican students in several schools in the Los Angeles metropolitan area; these students were tested and observed over a two-year period beginning in the spring of 1976. The purpose of the pilot study was to help in the development of a more refined research model and test battery for use in an extended study with a larger and more broadly-based sample.

The extended sample study, initiated in February, 1978, includes 400 third- and fourth-grade students in various sites in Los Angeles and in the Northeast. These students have been tested with a series of language proficiency measures in both English and Spanish. Analysis of the data should help identify those language skills and other factors that best predict student performance following transition from bilingual programs to classrooms in which English is the sole language of instruction. Based on this analysis and on subsequent follow up research, an index of factors to be considered in determining readiness for transition will be formulated, and materials will be developed for use by school personnel who must make instructional transition decisions. These materials will include a user's manual and guidelines for approaching the language assessment process from the standpoints of diagnosis and prescription. The diagnostic concern will be to determine which of the present capabilities and characteristics of students are conducive to successful transition to an English language instructional setting. The predictive concern will be to identify instructional remedies that will facilitate a successful transition.

Bilingual Evaluation Technical Assistance

Evaluations are the only source of information regarding students' academic progress in bilingual programs; they also serve as the basis for identifying exemplary projects. However, evaluations in bilingual education have varied in design and quality to the extent that they have been inadequate both for measuring program effect on student achievement and for identifying programs worthy of replication.

The reasons for the inadequacy of the evaluations include: (1) insufficient or contradictory specifications for evaluation by federal and state laws or regulations; (2) difficulty in describing the programs because of the vague nature of their goals and objectives; (3) insufficient training of program evaluators; (4) lack of appropriate instruments for testing bilingual students; (5) inappropriate program evaluation designs; and (6) delayed delivery of evaluation reports.

While our workshops will not be able to meet all these needs in bilingual evaluation, we at least hope to provide materials helpful in the following areas:

- the range of evaluation models suitable for use in various educational contexts
- the range of designs appropriate for meeting the information requirements of legislators
- the selection of an appropriate evaluation design that will meet decision needs at the program level and information requirements of various audiences
- the selection and/or development of suitable tests and other measures
- techniques for and considerations in the collection of data in bilingual education.

Workshops on the above kinds of concerns are intended to improve the quality and usefulness of evaluations by offering evaluation designs and techniques which users can adopt or adapt to meet local conditions or needs. The workshops are designed to meet the information needs of individuals at various decision-making levels. They will take into account legislative requirements under both federal and state statutes and the characteristics of bilingual education projects presently in operation.

The workshops are being developed in two series. The first series will include orientation modules presenting an overview of the evaluation process emphasizing the use of evaluation information for decision making. In addition to the orientation modules, the first series will contain information on the selection, development, and administration of instruments, including language assessment considerations, and appropriate evaluation designs for the analysis of data generated by the assessment process.

The second series will deal with procedures for documenting the instructional process, and evaluating program components such as staff training, materials development/acquisition/use, community involvement, and program management. Other modules may cover issues dealing with appropriate statistical techniques for analyzing the data, reporting the results to various audiences, interpreting the results by various audiences, and decision making based on evaluation findings.

Women and Minorities Evaluation Training

The target groups for this project are minority group members and women who are already working within schools or community settings where evaluation is either a primary or adjunct responsibility. The training we provide will thus be inservice, not pre-service. Our choice of clientele is based on the fact that legislation and regulation have motivated "equal employment" recruitment so that women and minorities have become much sought after entry-level employees. The progress of these groups in acquiring jobs needs to be supported by helping them to retain their positions or move on to more prestigious and responsible ones. The acquisition of skills related to program planning, program management, and program evaluation may be related to upward job mobility; in particular this appears to be an important factor for people working in social action programs of all types. Many federal and state educational and social action programs are currently staffed by women and minorities who have expertise in the subject areas of their responsibility. However, their job requirements may involve writing proposals, conducting needs assessments, arranging for program activities, and planning for or commissioning evaluations. Our work with educational agencies has shown us that many staff mem-

bers at lower administrative levels—whether or not minority group members—have needs for skills in these areas. Reports to us from previous participants in training sessions, workshops, and graduate courses given by CSE staff members suggest that skills in evaluation are and will continue to be highly marketable.

In sum, we have observed that in addition to education, social service endeavors in mental health, medicine and public health, drug treatment, housing and federal subsidy programs, and law enforcement are increasingly mandating evaluations. People filling leadership roles in such social service areas can benefit from practical evaluation training which even graduate programs in these areas do not now provide.

The project is being conducted in three phases. During the first phase, we will be concerned with answering such questions as:

- What jobs do women and minority members most commonly work in when evaluation is a major adjunct requirement? What is their potential mobility?
- Are the training needs of women and minorities the same?

Answers to these kinds of questions will determine the topics and scope of training, how to tailor the training to current expertise or job placement needs, and the options available for delivering the training.

During the second phase, in which materials will be developed, recruitment procedures will be formulated so as to reach the people who will benefit most from the training.

During the third phase, a small group of trainees will be chosen to participate in a materials tryout; based on these results, materials will be refined for future training purposes.

Evaluation Field Services

This unit provides training and technical assistance, on a cost-reimbursement basis, to evaluation practitioners, administrators, and teachers. This training is available to federal agencies, state and county departments of education, school districts, and individual schools. Our services draw upon CSE's expertise in the design and conduct of evaluations as well as the products, resources, and ideas developed and tried out in CSE's other programs and projects. The technical assistance or training may, where appropriate, include CSE workshops which are used as organizers for the training. These workshops include:

- **Evaluation Workshop I: Orientation to Evaluation**
- **Evaluation Workshop II: Needs Assessment**
- **Evaluation Workshop III: Program Planning**
- **Evaluation Workshop IV: Formative Evaluation**
- **Evaluation Workshop V: Summative Evaluation**
- **Teacher Appraisal for Improvement**

Services range from such broad evaluation topics as those listed above, to more specific concerns such as choosing the most appropriate evaluation design or selecting/developing measures to answer particular evaluation questions. We can combine and adapt our resources to meet the evaluation problems and needs of a particular user, rather than making the user fit into some predetermined set of activities.

Field Services reflects our traditional concern with evaluation training and related outreach activities. It is a vehicle for introducing new ideas and evaluation techniques, for expanding CSE contacts with educators with unique problems in evaluation, and for training graduate students and CSE interns. In addition, information gathered in the delivery of training and assistance in the actual conduct of evaluations generates feedback to our other program areas.

Teacher Appraisal for Improvement (TAFI) . . . a CSE workshop available through CSE Field Services or Phi Delta Kappa.

Dissemination

To disseminate our research through print, we have created and distributed different kinds of publications. Our first five evaluation monographs have been distributed to over 5,000 individuals. *Evaluation Comment* is available without cost to a readership of 10,000. Articles written by CSE staff and faculty associates may appear in the *CSE Report* series or the *CSE Working Paper* series; staff and faculty articles are also found in a wide variety of scholarly journals and in the ERIC system.

Non-print dissemination occurs through staff presentations at professional meetings; through annual CSE conferences and scholars invited to visit and work with us; through cooperation with other labs and centers; through testimony at legislative hearings. By responding to requests for training and technical assistance from schools, state and federal agencies, and legislative assistants, we have been able to infuse research-produced knowledge directly into policy deliberations.

As part of our concern with knowledge production and utilization, we are establishing contacts and dissemination arrangements with other agencies—LEA's, SEA's, and other R&D institutions. Our products and services are included in such sources as the Directory of California Dissemination Projects, being assembled by the Far West Laboratory, and the Human Resources Inventory, a resource provided to the Regional Information Exchanges. We are also exploring the possibility of making our workshops available to other educational agencies (for example, Phi Delta Kappa has already published our Teacher Appraisal for Improvement Workshop) in order to increase workshop dissemination, and at the same time to continue offering these and other resources through our own Field Services.

Additional CSE Activities

In addition to our on-going programs, CSE provides services to university, community, and government agencies. Examples include:

- Audit activities in relation to the evaluations of Experience-Based Career Education Programs
- Audit/evaluation of the Basic Skills Learning Centers established by South West Regional Laboratory and the Southeastern Educational Laboratory for the Community Services Administration
- Formative evaluation efforts in connection with Los Angeles Unified School District's desegregation efforts
- Follow-up studies deriving from CSE's evaluation of the California Early Childhood Education program
- Evaluation of teacher training activities for mainstreaming handicapped children sponsored by the California Regional Resource Center
- Evaluation of the Pediatric Intervention program for handicapped children conducted by the UCLA Medical School
- Analysis of writing skills of freshmen at selected university and college campuses
- Evaluation of in-service courses for the UCLA School of Engineering and the UCLA School of Medicine
- Evaluation of services provided to foreign students on the UCLA campus

Recent CSE Reports—1978

Skager, R. W. The Great Criterion-Referenced Test Myth. No. 95.

Alkin, M. C. The Title of This Speech is . . . . No. 96.

Alkin, M. C. A New Role for Evaluators. No. 97.

Wilcox, R. R. Some Comments on Selecting the Best of Several Binomial Populations or the Bivariate Normal Population Having the Largest Correlation Coefficient. No. 98.

Wilcox, R. R. Estimating True Score in the Compound Binomial Error Model. No. 99.

Wilcox, R. R. Estimating the Likelihood of False-Positive and False-Negative Decisions in Mastery Testing: An Empirical Bayes Approach. No. 100.

Wilcox, R. R. Applying Ranking and Selection Techniques to Determine the Length of a Mastery Test. No. 101.

Wilcox, R. R. Comparing Examinees to a Control. No. 102.

Walker, C. B. Standards for Evaluating Criterion-Referenced Tests. No. 103.

Burstein, L., Keesling, J. W., Conklin, J., & Doscher, M. L. Auditing a Large-Scale Evaluation: The Quality of Evaluative Information for the Assessment of Program Impact and for Decision Making. No. 104.

Burstein, L. Secondary Analysis: An Important Resource for Educational Research and Evaluation. No. 105.

Baker, E. L. Fun and Games: Their Relationship to Basic Skills in Elementary School Children. No. 106.

Dotseth, M., Hunter, R., & Walker, C. B. Survey of Test Selectors' Concerns and the Test Selection Process. No. 107.

Walker, C. B. Control Items: A Baseline Measure for Evaluating Achievement. No. 108.

Fischer, K. B. The Bilingual Prediction Project: Technical Report of the 1976 Pilot Study. No. 109.

Smith, L. S. Decisions and Dilemmas in Constructing Criterion-Referenced Tests: Some Questions and Issues. No. 110.

Conklin, J. Some Methodological Issues in Using Standardized Test Scores to Evaluate Large-Scale Educational Programs. No. 111.

Wilcox, R. R. Comparing Normal Populations to a Control. No. 112.

Porton, V. M. Grammatical Errors and Communication Breakdown. No. 113.

Millman, J. Determinants of Item Difficulty: A Preliminary Investigation. No. 114.

Wilcox, R. R. Estimating the Parameters of the Beta-Binominal Distribution. No. 115.

Malamuth, N. M., & Fitz-Gibbon, C. T. Tutoring and Social Psychology: A Theoretical Analysis. No. 116.

Fitz-Gibbon, C. T. An Analysis of the Literature on Cross-Age Tutoring. No. 117.

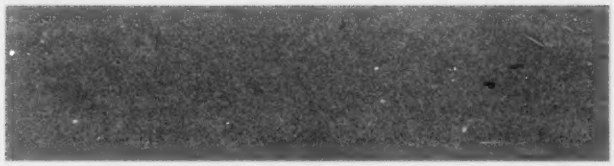
Fitz-Gibbon, C. T. A Survey of Tutoring Projects. No. 118.

Irizarry, R. Characteristics of Bilingual Programs Under Federal and State Legislative Policy: Implications for Program Development and Evaluation. No. 119.

Cabello, B., Elvenstar, D., Garcia, P., & Sisson, B. Variables Relevant to the Evaluation of Bilingual Education Programs: A Review of the Literature. No. 120.

Fitz-Gibbon, C. T. Tutoring: Some New Ideas. No. 121.

For information about ordering, contact Mark Young at the Center for the Study of Evaluation, UCLA Graduate School of Education, Los Angeles, CA 90024



Announcements of New CSE Publications

An Evaluation Kit . . . .

The *Program Evaluation Kit* contains eight books addressing topics most on the minds of people confronting the task of evaluating an educational program:

- The Evaluator's Handbook – A step-by-step guide
- How to Deal with Goals and Objectives
- How to Design a Program Evaluation
- How to Measure Program Implementation
- How to Measure Attitudes
- How to Measure Achievement
- How to Calculate Statistics
- How to Present an Evaluation Report

The Kit is designed to address the many questions that an evaluator might ask. Most published work in evaluation and educational research is heavy on theory and short on procedures that someone can actually use. This Kit answers questions procedurally, as though it were a consultant called in to help people do things now. Originally designed for practicing evaluators, the Kit's detailed advice has led to its use as a text for evaluation methods courses. Each of the Kit books can be used independently, without reference to any of the others.

The books, obtainable either as a Kit or individually, will be available in October, 1978 from Sage Publications, Inc., 275 S. Beverly Dr., Beverly Hills, CA 90210.

A Special Report on Bilingual Education . . . .

Characteristics of Bilingual Programs Under Federal and State Legislative Policy—Implications for Program Development and Evaluation by Ruddie Irizarry.

This report provides a review of federal and state bilingual education laws and an across-states summary which identifies the important variables to be considered in documenting and evaluating a bilingual program. Fourteen categories—such as population to be served, staff training requirements, student assessment, and evaluation reporting requirements—are used in organizing the summaries.

This report is available from the National Dissemination and Assessment Center—ESEA, Title VII, California State University, Los Angeles, School of Education, 5151 State University Drive, Los Angeles, CA 90032.

A Special Report on Cross-Age Tutoring . . . .

Setting up and Evaluating Tutoring Projects by Carol Taylor Fitz-Gibbon

Through a presentation of step-by-step planning questions, this report deals with the process of setting up and evaluating a tutoring project. The practical, systematic approach set forth in the report grows out of the author's experience as a teacher and out of her extensive research on cross-age tutoring as a compensatory strategy.

For ordering information, contact CSE Publication Services.

CSE COMMENTS

Three New CSE Evaluation Monographs . . .

#8 — Toward a Methodology of Naturalistic Inquiry in Educational Evaluation

In his monograph, E. Guba makes the case for naturalistic inquiry as distinguished from experimental inquiry. He then examines three major conceptual problems in doing evaluations using naturalistic methodology: boundary problems, focusing problems, and problems of authenticity. In so doing he summarizes work from other fields and demonstrates their applicability to evaluation.

#7 — The Logic of Evaluative Argument

The monograph by E. House takes the view that evaluation, lacking as it does the certainty of proof and conclusiveness, is fundamentally an act of persuasion. The analytical schema of the monograph grows out of a careful treatment of the premises of argument and agreement, distinguishing between the class of the "real"—facts and presumptions claiming validity vis-a-vis the universal audience, and the class of the "preferable" which includes values adhered to by a particular audience. Dr. House maintains that evaluation that is not mere oratory nor entirely arbitrary can greatly increase the possibilities for informed action.

#6 — Achievement Test Items—Methods of Study

This monograph examines methods of studying and characterizing certain types of achievement measures. The authors, C. Harris, A. Pearlman, and R. Wilcox, provide a model for describing student response data. They then review and critique the many measures of association and agreement which might be used for studying the equivalence and stability of achievement test items.

Other CSE Evaluation Publications

Educators continue to use the *CSE TEST EVALUATION SERIES*, a consumer's reference guide to standardized, published tests. Each volume contains a compendium of tests which have been evaluated for such characteristics as meaningfulness in relation to a specific goal, appropriateness for the intended student, usability for administration and decision making, and overall quality of standardization and technical excellence. The tests were judged by CSE staff trained to use a carefully devised evaluation procedure.

Each volume in the series focuses on a particular level of education for which a comprehensive list of educational goals has been developed by CSE. Superintendents, principals, testing officers, and other educators report that the CSE test evaluation books provide valuable assistance in selecting new test instruments and in the critical evaluation and revision of existing testing programs. The potential user need not have technical expertise in educational measurement and evaluation.

The following volumes are currently available:

CSE SECONDARY SCHOOL TEST EVALUATIONS
(1974—3 volumes, 900 pages, softbound 9" by 12")

CSE ELEMENTARY SCHOOL TEST EVALUATIONS
(1976—660 pages, softbound 9" by 12")

CSE-ECRC PRESCHOOL/KINDERGARTEN TEST EVALUATIONS
(1971—76 pages, softbound 9" by 12")

CSE-RBS TEST EVALUATIONS: TESTS FOR HIGHER-ORDER COGNITIVE, AFFECTIVE, AND INTERPERSONAL SKILLS
(1972—352 pages, softbound 9" by 12")

SYSTEM FOR OBJECTIVES BASED ASSESSMENT—READING (SOBAR)

SOBAR is a complete K-9 criterion-referenced system for evaluating reading instruction. The system contains a variety of components, including: catalogue of objectives, general manual, student profile, customized tests based on a bank of computerized test items, tests for each objective at each appropriate grade level cluster. Intended users of the system are administrators, teachers, curriculum directors, evaluators of reading programs, and students interested in reading skills.

SYSTEM OF OBJECTIVE-BASED EVALUATION IN READING—SPANISH (SOBER)

SOBER is a complete K-6 criterion-referenced system for evaluating reading instruction in the Spanish language. The system contains a variety of components including a catalogue of objectives, a user's manual, and a bank of items referenced to the objectives. K-3 materials are currently published by Science Research Associates, Inc.

For more information about the SOBAR or SOBER system write to Science Research Associates, Inc., 155 North Wacker Dr., Chicago, IL 60606.

Evaluation Field Services

CSE Workshops, Kits, and other resources from our research projects provide the bases for the training and technical assistance offered by Field Services. During the past year we have provided evaluation services to governmental agencies, school district, and schools throughout the United States. If you are interested in the range and costs of services available, contact:

James Burry, Director, Field Services
Center for the Study of Evaluation
UCLA Graduate School of Education
Los Angeles, CA 90024
(213) 825-4711

Evaluation Network is having its Fifth Annual Conference in Aspen, Colorado, August 14 - 16, 1978. For further information contact:

Professor Robert B. Ingle
School of Education
University of Wisconsin-Milwaukee
P.O. Box 413
Milwaukee, WI 53201
(414) 963-4729

Order Form

CSE Test Evaluation Series

Volume(s) of the CSE-ECRC Preschool/Kindergarten Test Evaluations ..... @ $5.00 each
Volume(s) of the CSE Elementary School Test Evaluations ..... @ $12.50 each
Set(s) of the CSE Secondary School Test Evaluations @ $22.00/set
Gr. 7&8 Gr. 9&10 Gr. 11&12 ..... or $8.00 volume.
Volume(s) of the CSE-RBS Test Evaluations: Tests for Higher-Order Cognitive, Affective, and Interpersonal Skills ..... @ $8.50 each

TOTAL $

CSE Monograph Series in Evaluation

Copy(s) of #1 — Domain-Referenced Curriculum Evaluation. W. Hively, et al., 1973 ..... @ $3.50 each
Copy(s) of #2 — National Priorities for Elementary Education. R. Hoepfner, et al., 1973 ..... @ $3.50 each
Copy(s) of #3 — Problems in Criterion-Referenced Measurement. Editors: C. Harris, M. Alkin, W. J. Popham, 1974 ..... @ $3.50 each
Copy(s) of #4 — Evaluation and Decision Making: The Title VII Experience. M. Alkin, et al., 1974 ..... @ $3.50 each
Copy(s) of #5 — Evaluation Study of the California State Preschool Program. R. Hoepfner & A. Fink, 1975 ..... @ $3.50 each
Copy(s) of #6 — Achievement Test Items: Methods of Study. C. Harris, A. Pearlman, & R. Wilcox, 1977 ..... @ $4.50 each
Copy(s) of #7 — The Logic of Evaluative Argument. E. House, 1977 ..... @ $4.50 each
Copy(s) of #8 — Toward a Methodology of Naturalistic Inquiry in Educational Evaluation. E. Guba, 1978 ..... @ $4.50 each

Total $

Send order to: Office of Publications
CSE - UCLA
Graduate School of Education
Los Angeles, CA 90024

☐ My Payment is Enclosed for This Amount

☐ Please Bill Me for This Amount

Add $.50 for Postage and Handling $ .50

Calif. Residents Add Sales Tax $

Grand Total $

CSE EVALUATION COMMENT UCLA

Center for the Study of Evaluation
UCLA Graduate School of Education
Los Angeles, California 90024

Non-Profit Org.
U.S. Postage
PAID
Los Angeles, Calif.
Permit No. 12378